Chapter 15
A Shop Bot for Web Market Intelligence

Miguel A. Morales-Arroyo
UNAM, Mexico

Foo Chee Yuan
Nanyang Technical University, Singapore

Lim Thian Muar
Nanyang Technical University, Singapore

Kwek Choon Hwee
Nanyang Technical University, Singapore

ABSTRACT

With the large amount of information available in the WWW, the ability to distinguish relevant from irrelevant data becomes a crucial factor. In this project, eight web scraping spiders were configured and evaluated for their functionality in order to determine their suitability for Interactive Digital media (IDM) start-ups to be utilized for competitive intelligence gathering. These spiders were chosen from the internet because of their availability and low cost. Each spider was configured and tested on two web sites. The evaluation process was first carried out individually to give a score to the spiders and then as a team to moderate the scores. The Web Info Extractor has the highest overall score as a web scraping spider while the Web Content Extractor has the best task analysis result. After the evaluation process, it is concluded that different spiders have varying capabilities and thus are suitable for different tasks. A spider that can handle more complex tasks is usually inherently more complex to configure and less-user friendly. Hence, in order to select the correct spider, companies should understand the tasks undertaken by their customers through basic task analysis as well as the knowledge of the amount of resources that they have at their disposal when it comes to configuring and operating the spiders.

INTRODUCTION

In today's business environment, the competitiveness of a company will depend on the ability of the enterprise to gather information from their respective business market and transfer that information into their strategic plans and decision making processes. In addition, an appropriate plan can then be made in response to changes in the market environment. In the past, information was gathered manually through a tedious process. But with the advancement of computing technology and the vast amount of information in the World Wide Web (WWW), gathering large amount of information is now possible by the introduction of web crawlers or scrappers.

Enterprises use information and knowledge to generate their product and/or services. Besides, the know-how needed to understand the market, costumers, providers, competitors, their competitive advantages and weakness. Enterprises need to constantly monitor its environment to keep competitive. They need to scrutinize potential threats and to seek for opportunities. Consequently, they require to access to external sources of information.

Market intelligence is a set of methodologies and technologies that are used for gathering, storing, analyzing, and providing access to data that enable the decision makers to make better business decisions (Keyes, 2006). The goal is to transform data into knowledge. Web market intelligence is a subset of business intelligence whereby data comes solely gathered from the web for analyzing. Market intelligence is the acquisition of environmental information that allows enterprises to keep a competitive edge. In other words, market intelligence provides an apparent advantage to business organizations on their competition by identifying the position, strategies of their competition, users' preferences and dislikes by gathering and analyzing external sources of information. Once information and knowledge have been acquired, it needs to be categorized, processed, prepared, and distributed for analysis to users inside the organization. The objective is to ensure that all the required valuable knowledge and information in relation to the competition and consumers' preferences are available as an input for the decision-making process (Carpe, 2007).

With the arrival of the Internet, new external sources of information and knowledge are available. Inherently, social interaction could be mined to gather consumers' preferences in relation to the own enterprise products and services or the competition. The same technologies used to share knowledge and information in the Internet can be utilized inside of the organization to empower staff to participate in the business intelligence process as collectors, analyst, and decision makers. The internet is considered as a competitive source of information (Brabston & McNamara, 1998) and seen as a cheap and quick mean to collect information (Wood, 2001). Scanning the internet environment offers a gateway to vast and varied information that can assist the enterprise in seeking and using information (Pawar & Sharda, 1997).

It has been found that companies are not exploiting the full potential of the internet (Dutta & Segev, 1999). Another study had showed that companies only use the internet as a means of portraying the corporate image (Adam et al., 2002). When it comes to use the Internet as a knowledge resource, there is a surprising low usage.

Information source identification and information gathering use numerous sources and techniques to obtain knowledge and information required by enterprises. Some sources of information are easy to accessed and free, such as financial statements in public companies, but others are expensive like marketing reports. Some not distributed information may be very difficult or almost impossible to obtain, for example competitors' strategies. Potential sources of unpublished information could be providers, consumers, or governmental representatives that have interaction with the competition and the regulatory process (Calof & Wright, 2008).

Related Content

Connecting Real and Virtual Neighbors: The Interplay between Physical Space, Civic Journalism, and Online Community

Robin Blom, Jonathan S. Morgan, Paul Zubeand Brian J. Bowe (2013). *International Journal of Interactive Communication Systems and Technologies (pp. 1-15).*

www.irma-international.org/article/connecting-real-and-virtual-neighbors/105653

Wearable Computers

Byron Havardand Megan Podsiad (2018). *Wearable Technologies: Concepts, Methodologies, Tools, and Applications (pp. 20-29).*

www.irma-international.org/chapter/wearable-computers/201951

Trustworthy Architecture for Wireless Body Sensor Network

G. R. Kanagachidambaresan (2018). *Wearable Technologies: Concepts, Methodologies, Tools, and Applications (pp. 333-362).*

www.irma-international.org/chapter/trustworthy-architecture-for-wireless-body-sensor-network/201967

An Interactive Platform for Sustainable Supply Chains

Ye-Sho Chen (2018). *International Journal of Interactive Communication Systems and Technologies (pp. 56-73).*

www.irma-international.org/article/an-interactive-platform-for-sustainable-supply-chains/214853

Software Reuse in Hypermedia Applications

Mario Mochicchioand Roberto Paiano (2002). *Interactive Multimedia Systems (pp. 195-203).*

www.irma-international.org/chapter/software-reuse-hypermedia-applications/24573